I024. A Deep-Learning Method for High-Throughput *FMR1* Triplet Repeat Screening

L. Ringel

Introduction: Approximately 1 in 200 females and 1 in 450 males are carriers of fragile X syndrome (FXS) in the US. High-throughput carrier screening requires an accurate and robust method for *FMR1* genotyping, which has been historically problematic due to the difficult-to-amplify CGG repeat that causes FXS in >99% of cases. Advances in PCR/capillary electrophoresis (CE) technologies have enabled the amplification and sizing of these triplet repeats within the clinically relevant range. However, existing CE trace interpretation requires trained manual operators or computational heuristic methods tailored to signal idiosyncrasies and peak morphologies. To address this issue, we developed a deep-learning algorithm that can automate reliable determination of the *FMR1* genotype. **Methods**: Blood and cell-line specimens were collected and processed with a commercially available test kit for detecting CGG expansions in *FMR1*. Multichannel CE relative fluorescence unit (RFU) values were pre-processed, sized, and scaled; candidate peaks above sample- and location-specific background thresholds were extracted for consideration. A multi-layered, multi-output convolutional neural network (CNN) classifier was developed to differentiate *FMR1* genotypic peaks from background noise. Repeat length was used to classify specimens into normal, intermediate, premutation, and full-mutation categories. The trained CNN was evaluated on an independent clinical validation cohort with reference *FMR1* genotypes assessed using a secondary PCR/CE technique and/or Southern blot analysis. **Results**: The CNN was trained on approximately 3,000 PCR/CE traces and the model was evaluated on an independent clinical cohort of >400 specimens. The pipeline took <120 minutes to train the CNN, and <10 minutes to process the validation cohort. The model achieved sensitivity and PPV of ≥98.5% at the genotype level, and ≥98.5% percent agreement at the categorical genotype level. Based on a Passing-Bablok regression analysis, there was no statistical difference between the repeat length reported by the CNN and the independent reference method. **Conclusions**: A deep-learning model accurately identified normal and expanded *FMR1* repeat lengths from CE electropherogram data. The model reliably distinguished legitimate peaks from artifacts and reliably genotyped complex residual clinical blood specimens. To our best knowledge, this system is the first deep-learning platform for automated PCR/CE analysis. It has the potential to expedite broader FXS screening and may also improve PCR/CE analysis of other repetitive, and non-repetitive, genetic markers.

ST010. Detection of Point Mutations in Paediatric Low Grade Glioma (PLGG) and Diffuse Intrinsic Pontine Glioma (DIPG) Patients: Validation of a Novel Liquid Biopsy Assay

M. Johnson

Introduction: Gliomas are the most common brain tumours in children. Paediatric low grade gliomas tend to have a good prognosis when a gross total resection is amenable. However, midline tumours are located in regions where resection is difficult, thus requiring a needle biopsy which frequently result in low tumour yields. Therefore, a minimally invasive method of molecular detection is critical to assist diagnosis, inform prognosis, and guide therapy. Previous reports on adult oncology indicate that liquid biopsy for brain tumours is challenging. Hence, we opted for a droplet digital PCR (ddPCR) based technology with pre-amplification to achieve maximal assay sensitivity. Here, we have tested a minimally invasive liquid biopsy tool for the detection of point mutations in cerebrospinal fluid (CSF) and plasma of paediatric brain tumour patients. **Methods:** CSF and blood were collected from patients harbouring the *BRAF* c.1799T>A (p.V600E) and *H3F3A* c.83A>T (p.K27M) point mutations. Samples were processed within 2 hours of collection. The QIAamp Circulating Nucleic Acid kit was used to extract circulating tumour DNA (ctDNA) from 2 ml of CSF and 3 ml of plasma. Primers and probes were designed and optimized for both mutations. Optimal conditions were 15 cycles of pre-amplification followed by ddPCR with 45 cycles, both at an annealing temperature of 54°C. **Results:** Linearity tests and mutation titrations confirmed the feasibility of detecting mutant droplets at low DNA inputs and low mutant allele frequencies (*BRAF* V600E 0.65 ng and 0.1% and *H3F3A* K27M 0.125 ng and 0.11%, respectively). We determined the background of the assay by running 24 wild-type samples for each mutation and calculated the Limit of Detection (LoD) and Limit of Blank (LoB) following RainDance's guidelines. For V600E the LoD and LoB were 15 and 8 droplets, respectively, and for K27M the LoD and LoB were 14 and 8 droplets, respectively. Out of 21 samples from V600E positive patients (16 plasma and 5 CSF), 4 out of 5 CSF samples tested positive by ddPCR. Out of 10 samples from K27M positive patients (8 plasma and 2 CSF), both CSF samples tested positive. This assay has an extremely low false positive rate which has resulted in excellent specificity (100% for *BRAF* V600E, 100% for *H3F3A* K27M) with very good sensitivity in CSF (80% for

BRAF V600E, 100% for *H3F3A* K27M). **Conclusions:** As demonstrated in this study, it is feasible to detect point mutations in the CSF of paediatric brain tumour patients at the time of diagnosis, potentially establishing a safer, less invasive diagnostic assay than the current standard fine-needle cranial biopsy. Due to the blood brain barrier, sufficient ctDNA is hindered from entering the bloodstream and consequently, is difficult to detect in plasma with our assay.

TT016. Utilization of Unique Molecular Barcodes in Next-Generation Sequencing Reveals Startling Differences in Library Chemistries in Read Depth and Allelic Frequencies

B. Anderson

Introduction: Clinical use of Illumina's TruSeq Custom Amplicon Myeloid next-generation sequencing (NGS) panel with unique molecular indexes (UMIs) and subsequent validation studies of Archer's VariantPlex panel revealed inherent problems in coverage depth in the Illumina libraries. Initial studies of the Illumina libraries showed varying allele frequencies (AF) among libraries, as well as variants present in one library but absent in another. Subsequent inclusion of UMI data confirmed the lack of unique coverage, as well as an inflation of true read depth. To provide quality patient care, four libraries were prepared and analyzed using three separate pipelines for all clinical samples. For validation, Archer libraries were prepared and analyzed in singlicate. AF of the synthetic control and previous samples were compared to assess coverage depth and library complexity. In addition, AF for all single nucleotide variants (SNVs) from 9 previous CAP samples were compared among 3 platforms: CAP digital-droplet PCR, Illumina, and Archer. **Methods**: For validation, data from 4 libraries and 3 pipelines from the Illumina assay were compared to the data from one Archer library and pipeline. Eighty previous and 36 parallel samples were evaluated for accuracy. Reproducibility was tested using 5 previous samples and 3 synthetic samples. Analytical sensitivity/Limit of Detection (LOD) studies were performed using 50 samples with AF = 10.5%: three synthetic samples combined into a single barcode and 47 samples with at least one variant present at AF = 10.5%. To assess for specificity, 12 negative samples were tested. **Results:** A total of 466 variants were tested: 392 >5.0%, 32 = 5.0%, 33 identified by visual inspection only in Illumina, and 9 not on the Archer panel. There was 100% precision on 48 variants >5.0% (CV = 11.55%) and four = 5.0% (CV = 21.2%). Analytical sensitivity/LOD studies with the Archer assay correctly identified 102 variants = 10.5%: 12 = 2.7%, 22 = 5%, 53 = 6%, and 15 variants not detected in the Illumina assay but confirmed by alternate methods. AF for 28 SNVs in 9 CAP samples were also compared. All SNVs were detected in the Archer assay with comparable AF. In the Illumina panel, three were not found in any of the four libraries, four were found in only one of the libraries, and three were found in two libraries. The AF were inconsistent and varied widely from the reported AF determined by CAP. **Conclusions:** Based on data from both panels, it is clear that UMIs are essential for measuring library complexity and for accurate and meaningful results in an NGS assay. In the Archer assay, the use of UMIs and unique start sites increases true coverage and library complexity, allowing for more accurate detection of variants with lower AF.